

基于改进 RT-DETR 的光伏电池可见光图像缺陷检测

陈志洪 范必双 陈梦圆 张 浩
(长沙理工大学人工智能学院 长沙 410114)

摘要: 为提升光伏电池缺陷检测在多类别、小目标等复杂场景下的准确性,基于 RT-DETR 提出了一种轻量化改进模型。首先,设计了一种轻量化的双路特征提取模块用于替代 ResNet 的基本模块,在减少模型参数的同时,使得模型具备局部与全局特征建模能力;其次,将基于注意力的同尺度交互模块(AIFI)中的多头注意力模块进行改良,删除部分注意力头,对剩下的注意力头进行线性变换,以缓解多头注意力机制中的冗余问题;最后,将网络中的卷积下采样模块替换为一种基于小波变换的下采样模块,改善下采样过程中存在的边缘信息丢失问题。实验结果表明,改进后的模型参数量相较原模型下降了 47.3%,计算量下降了 39.7%。在私有数据集上,改进模型的精确率、召回率、mAP@50 相较基准模型分别提高了 2.2%、4.1%、2.4%,这验证了改进模型的有效性;在公开数据集 PVEL-AD 上进行泛化性实验,改进模型的 mAP@50 达到了 70.6%,比基准模型高 6.3%,初步验证了其在未见数据上的泛化能力。

关键词: RT-DETR;光伏电池;缺陷检测;轻量化;注意力机制

中图分类号: TP391;TN36 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Defect detection of photovoltaic cells based on improved
RT-DETR for visible images

Chen Zhihong Fan Bishuang Chen Mengyuan Zhang Hao

(School of Artificial Intelligence, Changsha University of Science and Technology, Changsha 410114, China)

Abstract: To enhance the accuracy of photovoltaic cell defect detection in complex scenarios such as multiple categories and small targets, this paper proposes a lightweight improved model based on RT-DETR. Firstly, a lightweight dual-path feature extraction module was designed to replace the basic module of ResNet. This not only reduced the model parameters but also enabled the model to possess the ability of local and global feature modeling. Secondly, the multi-head attention module in (attention-based intra-scale feature interaction, AIFI) was improved by deleting some attention heads and performing linear transformations on the remaining attention heads to alleviate the redundancy problem in the multi-head attention mechanism. Finally, the convolutional downsampling module in the network was replaced by a downsampling module based on wavelet transformation to improve the problem of edge information loss during downsampling. The experimental results show that the number of parameters of the improved model has decreased by 47.3% compared to the original model, and the computational cost has decreased by 39.7%. On the private dataset, the improved model outperformed the baseline with increases of 2.2%, 4.1% and 2.4% in precision, recall and mAP@50, respectively, demonstrating its effectiveness. In addition, a generalization experiment was conducted on the public PVEL-AD dataset, where the improved model achieved an mAP@50 of 70.6%, which is 6.3% higher than the baseline model, providing initial evidence of its generalization ability on unseen data.

Keywords: RT-DETR; photovoltaic cell; defect detecting; lightweight; attention mechanism

0 引言

由于人们对煤炭、石油等不可再生能源的过度依赖,导致了日益严重的环境污染问题以及可能面临的能源短缺问题,广泛采用可再生能源技术对于脱碳和满足日益增长的

能源需求有着至关重要的意义。据统计,全球光伏系统于 2024 年的总装机容量达到了 2.2 TW,其中中国总装机容量超过了 1 TW,占全世界的 46% 左右^[1]。

由于光伏系统长期暴露于户外环境,易受热应力、湿度、机械冲击等多种因素影响,导致裂纹、热斑、遮挡等缺

陷,显著降低电池性能与发电效率^[2]。因此,高效、精准地在役缺陷检测已成为保障系统长期稳定运行的核心环节。

目前,对于光伏电池缺陷检测方法一般分为电学特性分析以及机器视觉方法两大类。电学特性分析主要通过分析 I - V 曲线等物理性能指标来评估电池状态。例如,朱显辉等^[3]采用二阶 Bézier 函数拟合 I / V 曲线,结合改进松鼠搜索算法实现参数辨识,从而实现对电池健康状态的精确表征。但是,此类方法往往侧重于宏观性能评估,难以实现对缺陷位置与形态的直观定位。近年来,视觉检测方法则凭借非接触、快速响应和易于集成等优势,在光伏组件表面缺陷检测中得到广泛应用。目前主流的视觉检测方法是基于深度学习的目标检测算法,按照是否显式生成候选区域,目标检测算法被分为双阶段检测算法和单阶段检测算法。

常见的双阶段检测算法有 R-CNN^[4]、Faster R-CNN^[5] 等,其主要流程包含候选框生成和目标分类及回归两个阶段。对于这类算法,研究者们主要从候选框生成机制和特征增强方面进行改进。例如,Zhang 等^[6]设计的多尺度特征区域候选融合网络(multi-feature region proposal fusion network, MF-RPN)通过多特征区域提议融合机制提升了预测框的准确性;伊欣同等^[7]在 Faster R-CNN 中引入残差通道注意力(residual channel attention, RCA),有效增强了模型在复杂背景下的特征提取能力;Su 等^[8]通过将互补注意力网络(complementary attention network, CAN)引入 RPN 构成区域建议注意力网络(region proposal attention network, RPAN),实现对候选框的精细调制,以提高检测精度。

常见的单阶段检测算法包括 SSD^[9]、YOLO^[10] 等,它们一般在特征图上进行密集预测,直接输出目标的类别和位置。近年来,多位学者针对光伏电池缺陷检测提出了多种解决方案。例如,杨丽等^[11]在 YOLOv8 的颈部网络中引入双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)与上下文聚合模块(context aggregation module, CAM),以提升网络对微小目标的检测能力;周颖等^[12]在 YOLOv8 中引入 CAM 模块以及多注意力检测头(multi-attention detection head, MADH),有效缓解了检测过程中存在的微小缺陷丢失、特征混淆等现象;Wang 等^[13]通过在 YOLOX 的特征提取阶段引入 SENet(squeeze-and-excitation network)注意力机制,并结合 ASFF(adaptive spatial feature fusion)自适应空间特征融合策略,有效增强了模型对多尺度特征的提取能力;陈俊生等^[14]采用 C2f-ST(CSPDarknet53 to 2-stage FPN-swin transformer)模块与 Enhanced SimAM 注意力机制对 YOLOv8n 进行改进,有效提升了模型抑制背景噪声的能力与对细小特征的提取能力;陈方彬等^[15]构建了 YOLO-MCSL 模型,该模型通过在 YOLOv8s 中引入 RT-DETR 中的 CCFF 跨尺度特征融合模块增强跨尺度特征融合,并采用 Focaler-SIoU 损失函数优化边界框回归,有效提升了模型在红外场景下对微小

缺陷的检测能力。

综上所述,深度学习于光伏电池缺陷检测有着良好的应用前景,且都取得了不错的效果,但现有研究所采用的方法多数依赖预定义的锚框与非极大值抑制(non-maximum suppression, NMS)^[16]等人工设计组件,这限制了模型端到端优化潜力。此外,现有研究主要集中在电致发光(electroluminescence, EL)和红外(infrared radiation, IR)图像。EL 图像能够揭示微裂纹等内部缺陷,但采集成本高,通常仅适用于生产环节;IR 图像能够利用温差定位异常区域,但难以区分缺陷类型和成因。相比之下,可见光图像获取方式简单、分辨率高,更适合大规模运维和户外巡检。然而,由于光照变化、复杂背景和缺陷尺寸较小,可见光图像的检测任务也具备着一定的挑战性。

基于此,本文选用近年来提出的端到端检测框架 RT-DETR^[17]对光伏电池可见光图像进行缺陷检测。RT-DETR 摒弃了锚框与 NMS,通过 Transformer^[18]结构实现全局特征建模,具备更强的端到端优化潜力。针对光伏可见光图像检测中存在的微小缺陷易丢失、特征易混淆等问题,本文在 RT-DETR 的基础上进行了改进,以提升模型在保持低参数数量的同时的检测精度与鲁棒性。通过在实际光伏图像数据集和高质量公开数据集上的实验测试,本文提出的改进 RT-DETR 模型在保持较低参数数量的同时,实现了更高的检测精度,为光伏系统的智能化检测和低成本运维提供了可行方案。

1 RT-DETR 模型

RT-DETR 模型是百度提出的一种基于 Transformer 的实时目标检测模型,其具体结构如图 1 所示。

作为一种端到端的单阶段目标检测器,RT-DETR 摒弃了传统目标检测中常用 NMS 后处理步骤。它通过输出固定数量的可学习查询向量(object queries),并采用匈牙利匹配的方法将预测框与真实框进行匹配,从而针对每个真实目标生成唯一的预测框。RT-DETR-r18 主要由 ResNet-18^[19]主干网络、高效混合编码器(efficient hybrid encoder)和解码器构成。RT-DETR 官方针对不同应用场景为模型提供了多种主干网络配置。本研究场景对模型推理延迟要求较高,因此选用计算开销较小的 ResNet-18 作为主干网络。高效混合编码器的主要组成为基于注意力的同尺度交互模块(attention-based intra-scale feature interaction, AIFI)与基于 CNN 的跨尺度特征融合模块(CNN-based cross-scale feature fusion, CCFF)构成。AIFI 本质为一个多头自注意力机制,为避免注意力机制带来的过大计算量和参数量,RT-DETR 只对主干网络最后一个阶段的输出特征图 S_5 进行注意力操作,得到增强后的全局上下文特征图 F_5 。CCFF 内的核心组件是一个基于 RepBlock^[20]的特征融合模块,RT-DETR 通过它对特征图 S_3 (主干网络倒数第 3 个阶段的输出特征图)、 S_4 (主干网

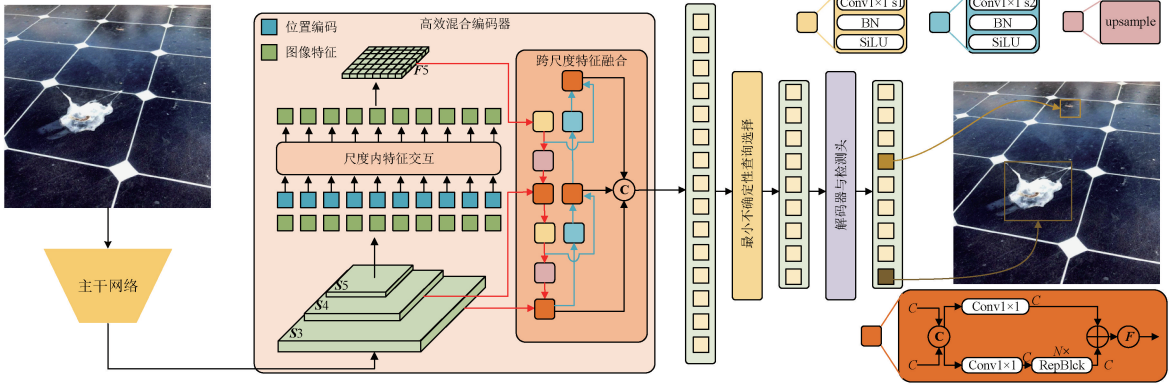


图 1 RT-DETR-r18 模型结构

Fig. 1 The structure of the RT-DETR-r18 model

络倒数第 2 个阶段的输出特征图)和 F_5 进行特征融合,以增强上下文感知能力。RT-DETR 使用 3 层 Transformer 解码器 300 个 object queries 与融合特征进行交互,逐步学习每个目标的位置与类别,实现端到端的目标检测流程。

2 改进 RT-DETR 模型

2.1 双路特征提取模块

光伏电池存在的缺陷复杂多样,在图像中一般呈现为尺

寸差异大的块状、条状以及一些不规则的图形。RT-DETR 模型的主干网络为 ResNet,它由多层卷积堆叠而成,具备良好的层级特征提取能力,但缺乏对全局长距离依赖关系建模的能力,因此对光伏电池可见光图像中的跨尺度缺陷不敏感。

为了解决这一问题,本研究构建了一种全局、局部建模的双路特征提取模块(dual path fusion block, DPFBLOCK)模块作为新的主干网络模块。DPFBLOCK 的具体结构如图 2 所示。

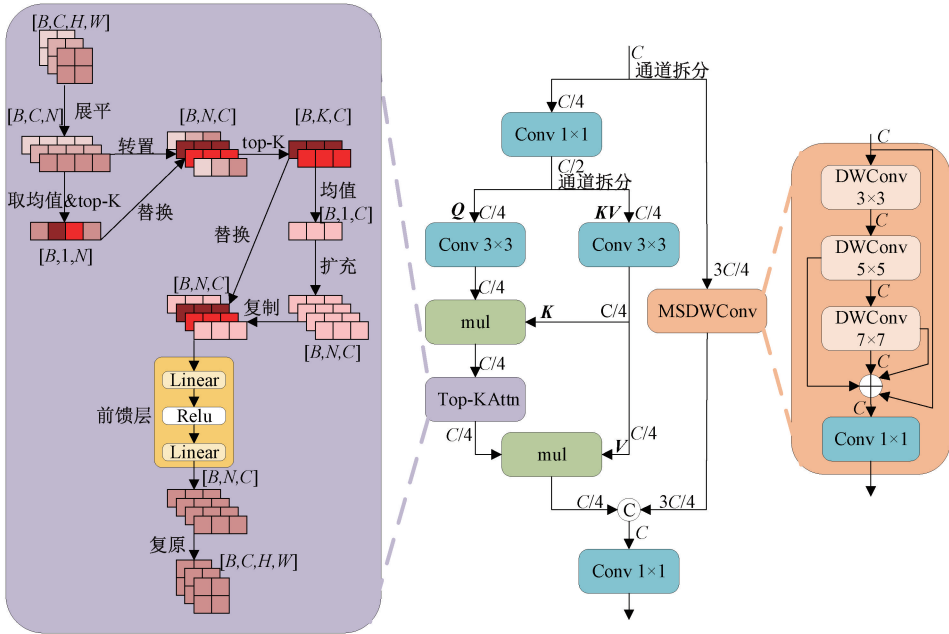


图 2 DPFBLOCK 结构

Fig. 2 Structure of the DPFBLOCK

结构上,该模块借鉴了 PConv^[21]的部分思想,将输入特征图在通道上拆分为两个功能互补的子分支,在通道数较少的分支上进行计算量大的全局建模,在通道数较多的分支上进行计算量相对较少的局部建模,最后通过一个 1×1 的卷积将两路特征进行信息整合。在全局建模分支

中,本研究采用了类 self-attention 机制,通过计算查询 Q 与键 K 之间的逐元素乘积获得每个空间位置的响应得分,并基于 Top-K 策略^[22]选取最具代表性的区域。具体而言,设定 $Q \odot K$ 的结果为 $X \in \mathbf{R}^{B \times C \times H \times W}$,按空间维展平得到:

$$\mathbf{X}_{\text{flat}} \in \mathbf{R}^{B \times C \times N}, N = H \times W \quad (1)$$

随后按通道维求平均得到各空间位置的重要性,公式如式(2)所示。

$$\mathbf{S}_s = \frac{1}{C} \sum_{c=1}^C \mathbf{X}_{\text{flat}}[b, c, n], \mathbf{S}_s \in \mathbf{R}^{B \times 1 \times N} \quad (2)$$

根据得分排序,获取前 K 个最具代表性空间位置的索引 T_K 。随后将 $\mathbf{X}_{\text{flat}}$ 转置并按索引 T_K 提取出显著特征。

$$\mathbf{X}^T = \text{permute}(\mathbf{X}_{\text{flat}}) \in \mathbf{R}^{B \times N \times C} \quad (3)$$

$$\mathbf{X}_K^T = \mathbf{X}^T[:, T_K, :] \in \mathbf{R}^{B \times K \times C} \quad (4)$$

对这些显著特征在空间维取均值得到通道重要性向量

$$\mathbf{S}_c = \frac{1}{K} \sum_{n \in T_K} \mathbf{X}^T[b, n, c], \mathbf{S}_c \in \mathbf{R}^{B \times 1 \times C} \quad (5)$$

将 $\mathbf{S}_c$ 扩展至 $\mathbf{R}^{B \times N \times C}$ 并在显著位置 T_K 上用 $\mathbf{X}_K^T$ 进行替换,形成增强特征 $\hat{\mathbf{X}}$,最后通过一个前馈神经网络进一步融合通道间的语义信息,并恢复特征到原始形状:

$$\mathbf{Y} = \text{permute}(\text{FFN}(\hat{\mathbf{X}})), \mathbf{Y} \in \mathbf{R}^{B \times C \times H \times W} \quad (6)$$

在全局分支中, K 值的选取直接影响了模型的计算开销和显存占用,为保证模型能够以较低的训练资源下实现一定的全局表征能力,本研究在 $0 \sim 0.3$ 的范围内对 K 值进行了实验,具体结果如表 1 所示。实验结果表明,当 K 取值在 $0 \sim 0.2$ 时,随着 K 的增加,模型的训练时间与显存消耗略微变大,但精度能够获得较为显著的提升;而当 K 进一步增大至 $0.2 \sim 0.3$ 时,训练时间与显存消耗均出现明显增长,但精度提升幅度相对有限。因此本文最终选定 $K=0.2$ 作为模型训练默认配置。

表 1 全局建模分支中 K 值选择的对比实验

Table 1 Comparative experiments on the selection of K values in the global modeling branch

K	Epoch Time/s	Memory/GB	mAP@50/%
0.05	91	3.23	43.2
0.10	92	3.25	45.6
0.15	95	3.26	47.2
0.20	101	3.27	47.5
0.25	119	3.29	47.3
0.30	133	3.31	48.1

在局部建模分支中,本研究提出了一种轻量化的多尺度特征提取模块 MSDWConv (multiple scale DWConv)。该模块通过多个不同感受野的 DWConv^[23] 提取多尺度特征,并与原始输入残差相加,最后使用 1×1 卷积完成通道融合。考虑到光伏电池可见光图像同一块区域可能混杂多种不同尺度和形态的缺陷,本文利用 DWConv 计算量小的特点,在局部分支并行建立了 1×1 、 3×3 、 5×5 的 DWConv 卷积,在保证低运算量的情况下实现多尺度特征的提取。

2.2 幻影多头注意力模块

多头注意力 (multi-head attention, MHA) 机制通过在多个子空间中并行建模上下文关系,从而学习到不同语义或空间下的注意分布。这种能力使得模型可以同时理解局部和全局等不同层级的信息。然而,Voita 等^[24]指出,在目标检测等空间聚焦任务下,多头注意力可能聚焦于图像中相似的空间区域和语义特征,存在一定的计算冗余。

AIFI 作为 RT-DETR 中高层语义信息交互的关键模块,其核心结构便是 MHA,因此该模块的输出中存在一定的注意力头冗余现象。即 MHA 的部分注意头所关注的区域高度重叠,这限制了模型在语义表达上的多样性和鲁棒性。为减轻这一现象,本研究借鉴了 GhostConv^[25] 的思想,对 MHA 进行改进,提出了幻影多头注意力模块 (ghost multi-head attention, GhostMHA) 模块。

具体而言,该模块在保留核心的注意力机制能力的基础上,减少 MHA 中显式计算的注意力头数量,并引入若干个低成本的线性变换层 (linear layer) 来生成额外的幻影注意力头 (ghost attention heads),从而在保持输出维度不变的情况下显著减少计算量。

在 MHA 中,假定输入特征 $\mathbf{X} \in \mathbf{R}^{B \times N \times C}$, $N = H \times W$ 将其进行线性映射得到:

$$\mathbf{Q} = \mathbf{XW}_q, \mathbf{K} = \mathbf{XW}_k, \mathbf{V} = \mathbf{XW}_v \quad (7)$$

并将 $\mathbf{Q}$ 、 $\mathbf{K}$ 、 $\mathbf{V}$ 按通道维划分为 H_r 个真实注意力头,其中每个头的通道数为 $\frac{C}{H_r}$,然后在这些头上进行注意力计算获得标准的注意力输出:

$$\mathbf{A}_r = \text{Softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{D}}\right)\mathbf{V} \quad (8)$$

随后为每个真实注意力头 H_{ri} 引入一个低成本的线性层 L_i ,以生成对应的幻影注意力:

$$\mathbf{G}_i = L_i(\mathbf{A}_{ri}), i = 1, 2, \dots, H_r \quad (9)$$

最后将 $\mathbf{A}_r$ 与 $\mathbf{G}_i$ 拼接,获得完整的注意力表示:

$$\mathbf{A} = \text{Concat}(\mathbf{A}_r, \mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_{H_r}) \in \mathbf{R}^{B \times N \times C} \quad (10)$$

最后经过线形层进行通道信息的整合于交互:

$$\mathbf{Y} = \mathbf{AW}_o \quad (11)$$

在该设计下,模型在保持整体输出维度不变的同时,有效降低了计算开销,并进一步增强了表示能力。与原始 AIFI 模块相比,替换标准 MHA 为 GhostMHA 后的新 AIFI 模块有效降低了参数数量和计算开销,并小幅度提高了模型的检测精度。GhostMHA 具体结构如图 3 所示。

2.3 小波下采样模块

在 RT-DETR 中,CCFF 模块负责整合不同尺度特征,在这个过程中,它通过步长为 2 的卷积层实现下采样操作,将不同尺寸的特征图变换到相同尺寸。然而,这种基于卷积的下采样方式在降低分辨率的同时,往往会丢失边缘、纹理等部分高频信息。为此,本文提出了一种基于 Haar 小波变换的小波下采样模块 (wavelet down sample,

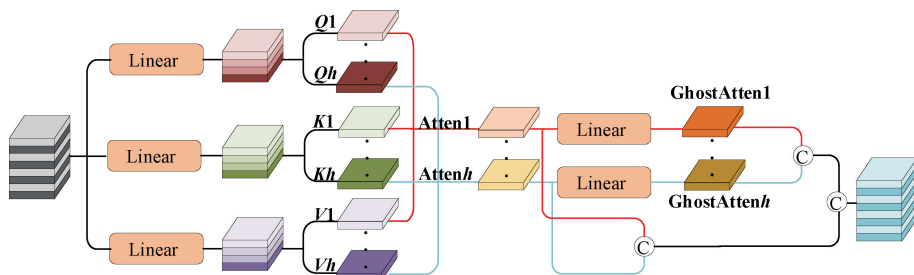


图 3 GhostMHA 结构

Fig. 3 Structure of the GhostMHA

WTDS)对 CCFF 中的卷积下采样层进行替换。Haar 小波结构简单、计算效率高,在保证下采样效率的同时能够较好地保留图像边缘与细节信息,非常适合实时检测场景。相比之下,其他常用小波基如 Daubechies、Symlets 或 Coiflets 在分解精度和对称性上更优,但滤波器长度较长,计算量较大,不利于轻量化实时检测任务,因此本研究最终选择 Haar 作为 WTDS 的小波基。

WTDS 是一种融合频域分解与可学习高通增强的下采样模块。设输入特征图为 $\mathbf{X} \in \mathbf{R}^{B \times C \times H \times W}$, 其经过 Haar 小波滤波器组 $\{\mathbf{h}_1, \mathbf{h}_b\}$ 进行二维分解, 获得特征图的低频子带 (LL) 与 3 个方向的高频子带 (LH, HL, HH):

$$\begin{cases} \mathbf{LL} = (\mathbf{X} * \mathbf{h}_l^T) * \mathbf{h}_l \\ \mathbf{LH} = (\mathbf{X} * \mathbf{h}_l^T) * \mathbf{h}_h \\ \mathbf{HL} = (\mathbf{X} * \mathbf{h}_h^T) * \mathbf{h}_l \\ \mathbf{HH} = (\mathbf{X} * \mathbf{h}_h^T) * \mathbf{h}_h \end{cases} \quad (12)$$

式中: * 表示卷积操作,输入特征经过步长为 2 的卷积后,输

$$\begin{cases} \tilde{\mathbf{L}}_{\text{LH}} = \text{DWConv}(\mathbf{LH}) \\ \tilde{\mathbf{L}}_{\text{HL}} = \text{DWConv}(\mathbf{HL}) \\ \tilde{\mathbf{L}}_{\text{HH}} = \text{DWConv}(\mathbf{HH}) \end{cases} \quad (13)$$

然后将经过滤波的高频子带与低频子带进行融合:

$$\mathbf{F} = \mathbf{L}\mathbf{L} + \tilde{\mathbf{L}}_{\text{LL}} + \tilde{\mathbf{L}}_{\text{HL}} + \tilde{\mathbf{L}}_{\text{HH}}, \mathbf{F} \in \mathbf{R}^{B \times C \times H/2 \times W/2} \quad (14)$$

最后通过 1×1 卷积进行通道间的信息交互, 得到最后输出的下采样特征:

$$\mathbf{Y} = \text{Conv}_{1 \times 1}(\mathbf{F}), \mathbf{Y} \in \mathbf{R}^{B \times C \times H/2 \times W/2} \quad (15)$$

相比传统卷积下采样方式,WTDS 能在降低分辨率的同时更有效地保留细节和边缘信息,提升网络对复杂纹理和小目标的感知能力。WTDS 的具体结构如图 4 所示。

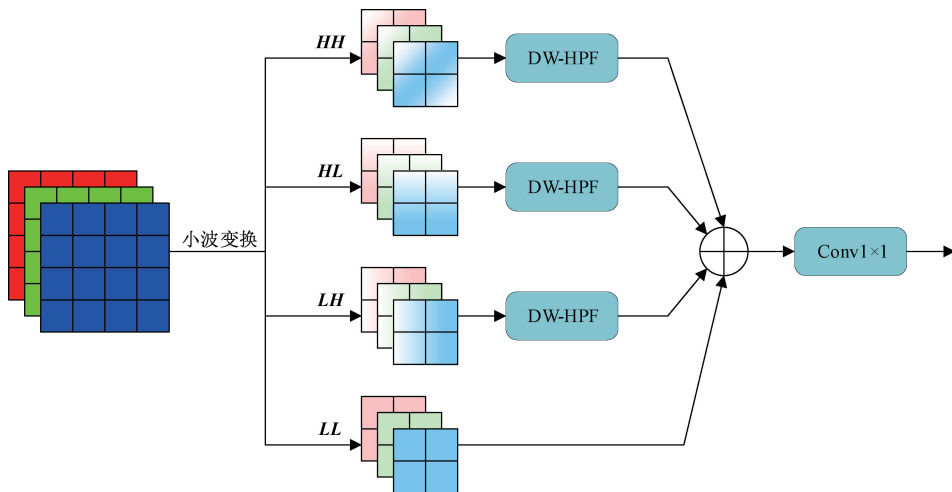


图 4 WTDS 结构

Fig. 4 Structure of the WTDS

3 实验结果及分析

3.1 实验配置

本文实验均在 PyTorch 2.0.1 框架下进行, CUDA 版

本为 11.8, 采用的显卡型号为 NVIDIA GeForce RTX 4090, CPU 型号为 Intel Xeon Platinum 8352V。

为确保实验结果的可复现性,本研究将随机种子统一固定为 0。模型训练基于 Ultralytics 框架,并采用 AdamW

优化器,动量因子设为 0.9,权重衰减系数为 1×10^{-4} 。初始学习率为 1×10^{-4} ,并按照余弦退火策略在训练末期衰减至初始值的 0.001 倍。

模型初始化采用 COCO 预训练权重。对于改进后的模型结构,由于部分层与原模型不完全兼容,仅加载可匹配参数,其余权重采用随机初始化。训练过程中未冻结主干网络,所有参数均参与反向传播更新。

损失函数各项权重设置如下:边界框回归损失权重为 7.5,分类损失权重为 0.5,DFL 损失权重为 1.5。

主要训练超参数配置如下:输入图像尺寸为 640×640 pixels,训练轮数为 250,批量大小为 4,优化器为

AdamW,初始学习率为 0.000 1,动量因子 0.9,权重衰减 0.000 1,数据加载线程数为 12。

3.2 数据集配置

实验数据由无人机采集获得,共获得 740 张光伏电池组件原始图像,涵盖 6 种类别:正常(139 张)、鸟粪(123 张)、积灰(181 张)、电气损伤(85 张)、物理损伤(82 张)和积雪(130 张)。部分样本示例如图 5 所示。数据集按类别以 7:2:1 的比例划分为训练集、验证集和测试集。为减轻小样本训练可能导致的过拟合问题,对训练集进行了数据增强操作,将训练集扩充至 1 800 张左右。增强方式包括随机翻转与旋转、随机裁剪、亮度调整、饱和度调整、色彩增强。

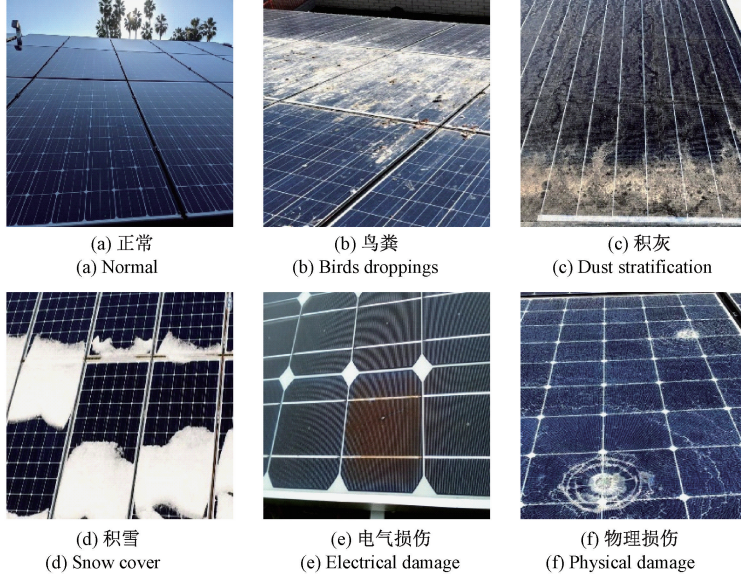


图 5 光伏电池缺陷检测数据集中包含的 6 类图像示例

Fig. 5 Examples of the six categories of images in the photovoltaic cell defect detection dataset

3.3 评价指标

本研究采用精确率(precision, P)、召回率(recal, R)、类平均精确率(mean average precision, mAP)、参数量(Params)、浮点运算数(GFLOPs)作为模型的评价指标。

P 的定义为模型预测为 1 的样本中真正为 1 的概率,它反映了检测结果的准确性; R 的定义为实际为 1 的样本在预测时为 1 的概率,它反映了模型的漏检率。两者的计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (16)$$

$$R = \frac{TP}{TP + FN} \quad (17)$$

式中:TP 为正样本中被预测为正的数,FP 为负样本中被预测为正的数,FN 为正样本中被预测为负的数量。

mAP 是对所有类别的平均精度(average precision, AP)取平均得到的指标,它表示着模型在所有类别上的整体检测性能。某个类别的 AP 指的是该类别的精确率-召回率(P - R)曲线下面积。AP 与 mAP 的计算公式如下:

$$AP = \int_0^1 P(R) dR \quad (18)$$

$$mAP = \frac{1}{N} \sum_{c=0}^N AP_c \quad (19)$$

式中: N 为总类别数量, c 为类别编号。

Params 为模型中所有可学习参数的总数量,用于衡量模型的大小。GFLOPs 为模型进行一次前向传播所进行的浮点运算数,单位为每秒 10 亿次,用于衡量模型的计算复杂度。

3.4 改进模块消融实验

为验证各模块对模型性能贡献,本文以 RT-DETR 为基准模型,在随机种子数设为 0 的条件下,于自建数据集上开展了系统性的消融实验,并分析了不同模块组合对性能指标的影响。具体结果如表 2 所示。

首先,将主干网络 ResNet-r18 替换为由 DPFBlock(记为模型 1)构成的轻量化主干网络,使模型参数量减少了约 41%,计算量降低了 38%,在大幅度降低模型复杂度的前提下,mAP@50 仅下降了 0.7%,仍维持了较高的检测精

表 2 改进模块消融实验结果

Table 2 Ablation results of the proposed module

基准模型	模型 1	模型 2	模型 3	$P/\%$	$R/\%$	mAP@50/%	mAP@50-95/%	Param/($\times 10^6$)	GFLOPs	FPS/fps
✓				57.1	48.5	48.2	20.1	20.08	58.2	93.4
✓	✓			56.8	48.8	47.5	19.4	11.51	35.9	82.6
✓		✓		58.4	53.1	50.1	22.2	19.78	57.2	101.0
✓			✓	63.1	49.1	48.8	20.8	18.85	55.0	97.1
✓	✓	✓		59.1	50.8	49.9	21.3	11.62	37.2	96.1
✓	✓		✓	61.2	49.2	49.5	21.0	10.68	35.0	92.7
✓		✓	✓	58.4	52.3	50.3	21.5	18.75	55.1	98.1
✓	✓	✓	✓	59.3	52.6	50.6	21.3	10.58	35.1	95.9

度,推理速度下降至 82.6 fps,但仍符合实时性要求;其次,将 RT-DETR 中 AIFI 中的多头注意力机制替换为 GhostMHA(记为模型 2),通过对真实注意力表示进行处理获得冗余注意力头表示,在略微降低运算量和参数量的情况下,使模型 P 提高了 1.3%、 R 提高了 4.6%,mAP@50 提升了 1.9%,同时推理速度提高到 101 fps;最后,通过替换 CCFF 模块中的下采样模块为 WTDS(记为模型 3),通过对输入图像进行小波变换,在下采样的同时保留了更

多图像细节,实现了模型参数量和计算量的小幅下降,同时 P 和 R 分别提升了 6% 和 0.6%,mAP@50 提升了 0.6%,推理速度提升至 97.1 fps。最后,将上述 3 个模块联合引入 RT-DETR,得到改进后的完整模型。

此外,为避免单次试验带来的偶然性并提升结果的统计可信度,本研究在随机种子数设为 0、123、531 这 3 种条件下进行了重复实验,并计算了不同模块下各主要性能指标(P 、 R 、mAP@50)的均值与标准差,具体结果如表 3 所示。

表 3 随机种子实验结果

Table 3 Random seed experiment results

(%)

基准模型	模型 1	模型 2	模型 3	$P(\text{mean} \pm \text{std})$	$R(\text{mean} \pm \text{std})$	mAP@50(mean $\pm$ std)
✓				57.3 $\pm$ 2.1	48.6 $\pm$ 1.1	47.3 $\pm$ 0.8
✓	✓			56.4 $\pm$ 1.7	49.4 $\pm$ 0.6	46.1 $\pm$ 0.8
✓		✓		58.2 $\pm$ 0.7	51.0 $\pm$ 1.8	49.0 $\pm$ 0.9
✓			✓	58.5 $\pm$ 5.1	49.3 $\pm$ 0.8	48.0 $\pm$ 2.6
✓	✓	✓	✓	59.1 $\pm$ 1.2	52.9 $\pm$ 0.5	50.1 $\pm$ 1.5

从表 3 均值与标准差可以看出,模型 1 下模型的性能维持了一个较好的精度,但存在一定的波动;模型 2 下模型的性能提升明显且波动小,这说明该模块在在不同初始化条件下均具有稳定的性能提升;模型 3 下模型的性能有所提升,但波动较大,对不同的初始化条件敏感。结合了 3 个模块的改进模型各项指标均取得最高均值,且标准差较低,表明各模块之间具有协同增益,性能提升并非偶发现象。

综合实验表明,改进后的 RT-DETR 模型在 P 、 R 和 mAP@50 上相比基准模型分别提高了 2.2%、4.1%、2.4%,且参数量下降了 47.3%,计算量下降了 39.7%,推理速度为 95.9 fps。这一结果验证了各个改进模块在提升模型精度与计算效率方面的有效性和互补性,改进后的 RT-DETR 结构如图 6 所示。此外,检测结果可视化对比如图 7 所示,可以观察到,原始模型存在一定的漏检、错检现象,而改进后的模型在目标边界的定位与类别识别上表现更为精准,进一步验证了本文所提出模块在提升模型实

际检测效果方面的有效性。

3.5 模型泛化性实验

为验证改进后 RT-DETR 模型在未见数据上的泛化能力,本文采用了由河北工业大学与北京航空航天大学联合发布的 PVEL-AD^[26]数据集进行泛化性实验。PVEL-AD 包含 36 543 张光伏电池电致发光图像,其中有 4 500 张图像为已标注图像,包含 12 类缺陷,主要为裂纹(crack)、断栅(finger)、黑芯(black core, bc)、粗线(thick line, tl)、星状裂纹(star crack, sck)、水平错位(horizontal dislocation, hd)、短路(short circuit, scc)等。为确保训练数据与泛化验证数据完全隔离,本研究从余下的 32 043 张图像中按索引均匀采样,每 600 张图像抽取 100 张图像,共抽取 5 000 张未见样本作为验证数据,并按照官方标注规则进行标注。泛化性实验结果如表 4 所示,与原始 RT-DETR 网络相比,改进后的网络在多个缺陷类别上的检测性能均有所提升。此外,原模型与改进模型在 PVEL-AD 上的可视化检测结果如图 8 所示。容易观察到,原模型存

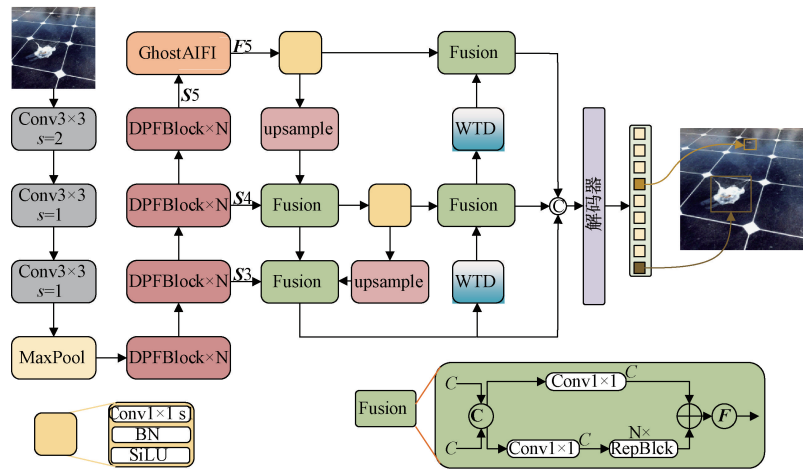
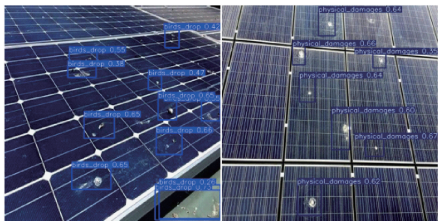


图 6 改进 RT-DETR 模型结构
Fig. 6 Improved RT-DETR architecture



(a) RT-DETR 检测效果
(a) RT-DETR detection results



(b) 改进 RT-DETR 检测效果
(b) Improved RT-DETR detection results

图 7 检测效果对比

Fig. 7 Detection results comparison

在一定的漏检现象,而改进后的模型能更加全面地识别图中的缺陷。综上结果表明,本文所提出的改进 RT-DETR 模型在对同场景下未见图片仍有较高的准确率和鲁棒性,一定程度上验证了模型的泛化性。

3.6 主流算法对比实验

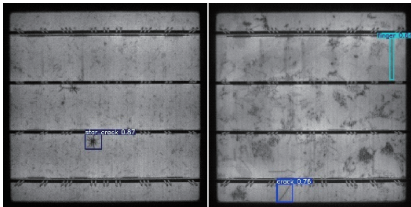
为评估改进后的 RT-DETR 模型在光伏电池缺陷检测上的性能,本文在私有数据集和训练环境下,将改进模型与主流目标检测模型(包括 Deformable-DETR^[27], DINO-DETR^[28]、YOLO 系模型等)进行了对比实验,对比指标包括 mAP@50、mAP@50-95 以及模型参数量,具体结果如表 5 所示。

从表 5 可以看出,本文提出的模型在精度与参数量之间取得了较好的平衡,在 10.58×10^6 参数量的条件下,本方法取得了 50.6% 的 mAP@50 和 21.3% 的 mAP@50-95,分别较原始模型提升了 2.4% 和 1.2%。与其他 Transformer 类检测器的轻量化版本相比,本研究的改进模型以更少的参数,实现了相对较高的检测精度。对于

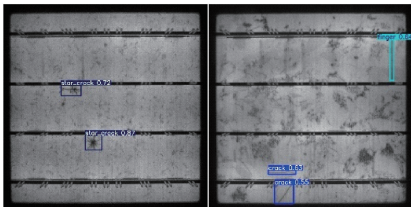
表 4 泛化性实验结果

Table 4 Generalization experiment results

模型	mAP@50/%							
	all	crack	finger	bc	tl	sck	hd	scc
RT-DETR	64.3	71.4	97.3	98.7	80.1	62.8	98.4	91.9
本研究	70.6	73.3	97.2	98.8	77.0	74.2	99.5	93.1



(a) RT-DETR 检测效果
(a) RT-DETR detection results



(b) 改进 RT-DETR 检测效果
(b) Improved RT-DETR detection results

图 8 泛化性实验可视化结果对比

Fig. 8 Comparison of visualization results from generalization experiments

表 5 主流模型对比实验结果

Table 5 Experimental results of comparisons with mainstream models			
模型	mAP@50/ %	mAP@50-95/ %	Param/ ($\times 10^6$)
本研究	50.6	21.3	10.58
RT-DETR-r18	48.2	20.1	20.08
Deformable-DETR-r18	49.7	21.6	23.38
DINO-DETR-r18	29.5	12.0	30.16
YOLOv5n	40.2	14.4	2.60
YOLOv5s	43.0	18.9	9.10
YOLOv8n	48.1	23.1	3.50
YOLOv8s	46.5	22.4	11.40
YOLO11n	50.0	22.7	2.60
YOLO11s	43.2	20.5	9.40

YOLO 系列模型, 尽管它们的参数量更小, 但在 mAP@50 指标上均低于改进模型。综上所述, 文中提出的改进模型在光伏电池缺陷检测任务中保持轻量化架构的同时兼顾了一定的精度, 具备较好的检测性能和良好的应用前景。

4 结 论

针对光伏电池可见光图像中缺陷种类繁多、尺度差异显著等问题, 本文提出了一种改进的轻量级 RT-DETR 模型。具体而言, 该模型通过将主干网络替换为 DPFBLOCK 构成的轻量化主干网络, 在降低参数数量的同时, 保留了一定的多尺度特征提取能力与全局建模能力; 通过将 AIFI 中的多头注意力替换为 GhostMHA, 缓解了传统多头注意力存在的特征冗余的问题并提升了模型的检测性能; 通过将 Fusion 前的卷积下采样改为 WTDS 模块, 增强了下采样过程中边缘特征的保留, 有效提高了模型的检测精度。为验证模型改进的稳定性, 本研究在 3 个随机种子下进行了重复实验。根据实验结果表明, 改进后的模型性能在不同随机初始化下均有所提升, 且波动不大。模型检测结果的可视化分析进一步验证了改进模型在复杂背景下的鲁棒性与对不同尺寸缺陷的敏感性。

在未来的工作中, 研究者将致力对改进 RT-DETR 模型进行进一步优化, 通过引入轻量化模块提升计算效率, 以适应高强度工业环境下的实时检测需求。同时, 将探索同一时间下光伏电池的红外与可见光图像的融合, 并送入改进模型进行训练, 从而实现光伏电池缺陷的全方位检测与精确诊断。

参考文献

[1] International Energy Agency. Snapshot of global PV markets 2025[R]. Paris: IEA PVPS, 2025.

[2] 刘玉洪, 吴一全. 基于机器视觉的太阳能电池片缺陷检测算法综述[J]. 光学精密工程, 2024, 32(6):

868-900.

LIU Y Q, WU Y Q. Review of defect detection algorithms for solar cells based on machine vision[J]. Optics and Precision Engineering, 2024, 32(6): 868-900.

[3] 朱显辉, 崔世炜, 鲁双峰. Bézier 函数协同改进松鼠搜索算法共同优化的光伏电池参数辨识[J]. 电子测量与仪器学报, 2024, 38(10): 191-200.

ZHU X H, CUI SH W, LU SH F. Photovoltaic cell parameter estimation through collaborative optimization of the Bézier function and the improved squirrel search algorithm[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(10): 191-200.

[4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[5] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[6] ZHANG X, HOU T, HAO Y W, et al. Surface defect detection of solar cells based on multiscale region proposal fusion network[J]. IEEE Access, 2021, 9: 62093-62101.

[7] 伊欣同, 单亚峰. 基于改进 Faster R-CNN 的光伏电池内部缺陷检测[J]. 电子测量与仪器学报, 2021, 35(1): 40-47.

YI X T, SHAN Y F. Photovoltaic cell internal defect detection based on improved Faster R-CNN [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(1): 40-47.

[8] SU B Y, CHEN H Y, CHEN P, et al. Deep learning-based solar-cell manufacturing defect detection with complementary attention network [J]. IEEE Transactions on Industrial Informatics, 2020, 17(6): 4084-4095.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]//European Conference on Computer Vision, 2016: 21-37.

[10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[11] 杨丽, 杨晨晨, 杨耿煌, 等. 改进 YOLOv8 的光伏电池缺陷检测算法[J]. 电子测量技术, 2025, 48(1):

- 92-99.
- YANG L, YANG CH CH, YANG G H, et al. Improved photovoltaic cell defect detection for YOLOv8[J]. Electronic Measurement Technology, 2025, 48(1): 92-99.
- [12] 周颖, 颜毓泽, 陈海永, 等. 基于改进 YOLOv8 的光伏电池缺陷检测[J]. 激光与光电子学进展, 2024, 61(8): 245-255.
- ZHOU Y, YAN Y Z, CHEN H Y, et al. Defect detection of photovoltaic cells based on improved YOLOv8 [J]. Laser & Optoelectronics Progress, 2024, 61(8): 245-255.
- [13] WANG J J, BI L, MA X D, et al. An efficient YOLOX-based method for photovoltaic cell defect detection[J]. Instrumentation, 2024, 11(2): 83-94.
- [14] 陈俊生, 陈沂蒙, 刘明杰, 等. 基于 MFES-YOLOV8n 的光伏电池缺陷检测方法[J]. 仪器仪表学报, 2025, 46(6): 251-262.
- CHEN J SH, CHEN Y M, LIU M J, et al. A defect detection method for photovoltaic cells based on MFES-YOLOV8n[J]. Chinese Journal of Scientific Instrument, 2025, 46(6): 251-262.
- [15] 陈方彬, 赵仲勇, 王建, 等. 基于 YOLO-MCSL 的轻量化智能电表热缺陷目标检测方法[J]. 仪器仪表学报, 2025, 46(8): 108-119.
- CHEN F B, ZHAO ZH Y, WANG J, et al. A lightweight thermal defect detection method for smart electricity meters based on YOLO-MCSL[J]. Chinese Journal of Scientific Instrument, 2025, 46(8): 108-119.
- [16] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]// European Conference on Computer Vision, 2020: 213-229.
- [17] ZHAO Y A, LYU W Y, XU S L, et al. Detrs beat yolos on real-time object detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv preprint arXiv: 1706.03762, 2017.
- [19] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [20] DING X H, ZHANG X Y, MA N N, et al. Repvgg: Making vgg-style convnets great again [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13733-13742.
- [21] LIU G L, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions [C]// European Conference on Computer Vision (ECCV), 2018: 85-100.
- [22] LU W, YANG X, CHEN S B, et al. Lwganet: A lightweight group attention backbone for remote sensing visual tasks [J]. ArXiv preprint arXiv: 2501.10040, 2025.
- [23] ANDREW G, ZHU M L. Efficient convolutional neural networks for mobile vision applications [J]. Mobilenets, 2017, 10(2): 151.
- [24] VOITA E, TALBOT D, MOISEEV F, et al. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, therest can be pruned[J]. ArXiv preprint arXiv:1905.09418, 2019.
- [25] HAN K, WANG Y H, TIAN Q, et al. Ghostnet: More features from cheap operations[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [26] SU B Y, ZHOU Z, CHEN H Y. PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection [J]. IEEE Transactions on Industrial Informatics, 2022, 19(1): 404-413.
- [27] ZHU X ZH, SU W J, LU L W, et al. Deformable detr: Deformable transformers for end-to-end object detection[J]. ArXiv preprint arXiv:2010.04159, 2020.
- [28] ZHANG H, LI F, LIU SH L, et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection[J]. ArXiv preprint arXiv:2203.03605, 2022.

作者简介

陈志洪, 硕士研究生, 主要研究方向为深度学习、缺陷检测。

E-mail:1539712446@qq.com

范必双(通信作者), 博士, 教授, 博士生导师, 主要研究方向为机器人及人工智能、电力电子系统控制及电机驱动、电能质量及配电网柔性消弧控制、新能源发电及其灾害防治。

E-mail:fanbishuang@qq.com

陈梦圆, 硕士研究生, 主要研究方向为深度学习。

E-mail:1992675344@qq.com

张浩, 硕士研究生, 主要研究方向为图像处理。

E-mail:1159657605@qq.com